Executive Summary - Analysis of Olympic Athletes' Build

Ethan Liu and Vaughn Hajra

Background

Since Ancient Greece, Olympians have been regarded as real-life superheroes. Competing in feats of strength, speed, and teamwork, Olympians have to be in peak physical condition for their sport. In this study, using a robust dataset that spans over one hundred years of Olympics, we investigate changes in the physical builds of Olympic athletes over time.

Inspired by existing research papers that investigate the optimal height of olympic swimmers, trends in the USA Olympic Gymnastics' height, and the overall stagnation of height in the US, we begin by first doing some exploratory data analysis. Then we use Vector Autoregression to look at trends in height and weight for Olympic athletes. Finally, we employ both multiple linear regression and Support Vector Regression to find which athletes are ahead of their time, based on their physical build.

Methods

- ***Vector Autoregression (VAR)*** - Vector autoregression is a kind of multivariate time series that is used whenever multiple time series influence each other. It is used both for forecasting, as well as examining causality and fluctuations based on the relationship between several variables. We discuss the requirements for usage of VAR, and then use it for forecasting, Granger causality, and Impulse Response analysis.
- ***Support Vector Machines (SVM)*** - Support Vector Machines work by dividing different groups of data while maximizing the margin between the two. We discuss how SVM works, the Kernel trick, and more. We are technically using Support Vector Regression for this project, which functions similarly to SVM but we have a continuous response variable. Challenges of SVM for this project include time complexity and computing constraints, but the benefits are a more precise model that outperforms linear regression.

Results & Suggestions

First, we find that male height stopped increasing for Olympic athletes in the 1970s. We interestingly find that since then, female height has continued to increase, and potentially stagnated only recently (but may still be increasing). This may be because of disparities in investment into male and female sports on the global scale, and the unlocked potential of female athletes. Weight for both male and female Olympic athletes has continued to increase, even as height has stagnated, and this weight is likely muscle mass. This is likely because of new training methods and science-based training that allow athletes to build more muscle.

When conducting regression (using both linear regression and support vector machines) we next attempted to predict the year in which an athlete competed using their biological data. We were able to do this fairly accurately with a MSE of ~18 years, or 4.5 Olympics. We furthered this by investigating athletes with large residuals, and looking into why their height and weight stood out. This analysis shows that there are multiple eras of Olympic athletes, and future studies investigating these time periods (potentially paired with global politics) would be very interesting.